DOCUMENT RESUME

ED 246 059 TM 840 293

AUTHOR Williams, David
TITLE Naturalistic Evaluation--One Case. Preliminary Draft.
PUB DATE Apr 84
NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Case Studies; Ethnography; *Evaluation Methods; Research Problems; *Standards
IDENTIFIERS *Evaluation Problems; Naturalistic Observation; Participant Observation; *Standards for Evaluation Educ Prog Proj Materials

ABSTRACT

The purpose of this paper is to summarize some of the author's logical and empirical reflections on the practice of using naturalistic techniques to conduct evaluations. The main assertion to be supported by these reflections is that naturalistic evaluation is possible and in some cases powerful. This assertion is supported through a review of (1) what naturalistic inquiry is, (2) criteria for judging the quality of a naturalistic study, (3) the purposes of evaluation, and (4) the "Standards for Evaluations of Educational Programs, Projects and Materials" developed by the Joint Committee on Standards for Educational Evaluation. The criteria for judging naturalistic inquiries are juxtaposed with the evaluation standards to identify criteria from the two perspectives which both support and conflict with one another. Finally, the author's experience conducting a naturalistic evaluation is analyzed in terms of those criteria which appear to conflict to see how the logical conflicts match experiential conflicts. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED246059

NATURALISTIC EVALUATION -- ONE CASE

David Williams

Brigham Young University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. D. Williams

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 293

To be presented at the annual meeting of the American Educational Research Association. New Orleans, Louisiana, April, 1984.

TABLE OF CONTENTS

Perspective..... 2

Logical Comparisons..... 4

Evaluation Standards..... 5
Utility Standards..... 5
Feasibility Standards..... 6
Propriety Standards..... 6
Accuarcy Standards..... 7

Criteria for Judging Naturalistic Inquiries..... 8

Comparison of Standards and Criteria.....12
Conflicts.....13
Summary of Logical Comparisons.....18

Experiential Analysis.....19

One Case of Naturalistic Evaluation.....19
Context.....19
Evaluation Procedures.....21

Analysis of Conflicts Between Criteria and Standards in Practice.....23
Analysis of this case in Light of Problems identified in the Literature.....28
Summary of Experiential Analysis.....29

Conclusions.....30

Table 1 - Comparison of Evaluation Standards and Criteria for Judging Naturalistic Inquiry.....31

Table 2 - Summary of Which Potential Conflicts and Which Criteria and Standards were Emphasized in One Case of Naturalistic Evaluation.....32

References.....33

Preliminary Draft - March, 1984

NATURALISTIC EVALUATION -- ONE CASE

David Williams
Brigham Young University

"You can't evaluate our course without participating in it, either as a student or as a teacher."

That statement in the initial evaluation orientation meeting with instructors of a youth leadership class was my first clue that I would soon be doing a naturalistic evaluation. As a university evaluation service, we had conducted a series of course evaluations for the Physical Education College. Most of those studies had used survey and quasi-experimental designs to help faculty and adminsitrators assess the value of the courses and to provide recommendations for improving them.

Having recently completed a study on the practices of naturalistic researchers, I interpreted this instructor's warning as an opportunity to explore the potential of the naturalistic paradigm in a course evaluation context. That experience led to a logical comparison of the ideals of the two forms of inquiry: naturalistic and evaluative.

The purpose of this paper is to summarize some of those logical and empirical reflections on the practice of using naturalistic techniques to conduct evaluations.

The main assertion to be supported by these reflections is that naturalistic evaluation is possible and in some cases powerful. Logical weaknesses will not necessarily be weaknesses in practice and practical strengths may not be logically predictable. This is so because actual evaluation procedures are products of negotiation between evaluators, producers of the evaluand and users of the evaluand. The ideals of each of these parties must be tempered to facilitate the negotiation. Therefore, how "naturalistic" the naturalistic methods used actually are and how ideally "evaluative" the study ends up being depends immensely on the actors and their expectations, purposes, abilities, resources, etc.

This assertion will be supported through a review of 1) what naturalistic inquiry is, 2) criteria for judging the quality of a naturalistic study, 3) the purposes of evaluation and 4) the Standards for Evaluations of Educational Programs, Projects and Materials developed by the Joint Committee on Standards for Educational Evaluation.

The criteria for judging naturalistic inquiries will be juxtaposed with the evaluation standards to identify criteria from the two perspectives which both support and conflict with

one another. Finally, my experience conducting a naturalistic evaluation will be analyzed in terms of those criteria which appear to conflict to see how the logical conflicts match experiential conflicts.

The relevance of this report is based on the assumption that a more realistic, data-based understanding of the strengths and weaknesses of using naturalistic inquiry to evaluate courses, programs, projects, etc., will encourage evaluators and evaluation consumers to use this method effectively when it is most appropriate.

Before introducing the criteria for judging naturalistic inquiry and evaluations, a brief literature review and perspective set the stage for the logical and experiential analyses.

Perspective

The purposes and procedures of naturalistic inquiry and educational evaluation traditionally have been divergent. Identified by a variety of labels (qualitative research, ethnography, participant observation, field work, phenomenology, etc.), naturalistic inquiry has evolved through anthropology, sociology, and other disciplines as a means of "understanding" people and the meaning behind their activities. The focus has been on describing human processes and using participant informers' views as guides to generating hypotheses and developing theories about those processes. Naturalistic inquirers usually view the people under study as collaborators in this discovery process. The inquiry procedures are non-judgmental, with relatively less emphasis on outcomes or products then on processes.

Educational evaluation has evolved more recently through government requirements and the experimental research paradigm as a means of estimating the merit or worth of educational processes and products. The focus has been on defining the criteria of value for a given study, collecting data on the activities of interest, and comparing activities and achievements to the criteria to assess the value of the evaluand. Although clear descriptions of the products and processes to be evaluated are essential, the emphasis has been on using those descriptions to make value judgments rather than to "understand" participants' meanings or to generate hypotheses and theories to explain participants' actions. Generally, the results of evaluations are intended for immediate practical use by policy makers, product developers, consumers, etc.

In spite of these differences in purpose, evaluation theorists (Stake, 1975; Guba, 1978, Guba & Lincoln, 1981; Patten, 1980) have suggested that the naturalistic approach be used as an alternative paradigm for guiding evaluators. They juxtapose

the techniques of ethnographers and participant observers with those of experimenters, expert witnesses, opinion pollers and other empiricists to argue that the qualitative perspective, defensibly obtained, can radically improve the evaluation enterprise.

In contrast, Wolcott (1982) has warned that although the naturalistic perspective clearly has merit for evaluators, the implications of adapting techniques from anthropology, sociology, and other fields as new evaluation "tools" are not clear. He suggests that an ethnographic orientation and the use of ethnographic collection procedures may be helpful to evaluators; but they should not presume to be doing anthropological ethnography. In other words, he claims the criteria for a successful ethnography (one form of naturalistic inquiry) may not be entirely compatible with criteria for a successful evaluation.

In addition to theorists, a few researcher-evaluators from sociology, anthropology, educational evaluation and other fields have described problems they encountered when using naturalistic methods in studies intended to inform policy makers and help audiences reach evaluative conclusions (see for example: Fetterman, 1982, 1983; Smith and Robbins, 1982; Stake and Easley, 1978; Williams, 1981). Although most of these reports (which involved studies at multiple sites, and large budgets) focused on the problems of doing multiple case studies in evaluative contexts, they also identified a variety of problem areas which are applicable to more common small-scale individual naturalistic evaluations:

1. The importance of timeliness in evaluation may conflict with the need for long-term engagement on site in naturalistic inquiry.

2. The emphasis on the emic perspective in naturalistic studies frequently but not always contrasts with the emphasis on etic or evaluator/decision-maker identified issues in evaluations.

3. The naturalistic policy of withholding judgment and stating biases explicitly has sometimes conflicted with the expectation that the evaluator will make evaluative judgments and recommendations.

4. The naturalistic inquirers' effort to provide a realistic and contextually rich picture in which the phenomena of interest are studied may distract from the evaluators' emphasis on a relatively small set of evaluation objectives.

5. The evaluation sponsors' typical expectation that quantifiable instruments, created <u>a priori</u>, will be used could interfere with the naturalistic inquirers' preference to use qualitative

observations and interviews either exclusively or preparatory to creating instruments.

6. A common view by program participants that evaluators are looking for problems and deficiencies in the program may block the naturalistic inquirers' efforts to develop roles in which they can discover how systems work from the insiders' perspective.

7. The need to develop rapport with all informants in a naturalistic study is complicated in an evaluative study in which certain informants (sponsors of the study usually) have financial power over the life of the project and other informants have power over the degree to which they will give their insider perspective to the inquirer. The evaluator must mediate between these conflicting roles and interests.

8. The naturalistic orientation that the purpose of inquiry is to understand may conflict with the evaluation (particularly formative evaluation) emphasis on describing, judging, and then improving.

9. The naturalistic policy of allowing issues to emerge and analyses to proceed inductively conflicts (except in uncommon goal-free evaluations) with the common practice of evaluators to identify the objectives of the evaluand and then formulate evaluation objectives in a proposal before the study ever begins.

10. Naturalistic reports and data records tend to be long, detailed and descriptive, leaving judgments to the reader. Evaluation audiences typically want to know "the bottom line" and problems arise for the naturalistic evaluator who must summarize results to a relatively short length for most readers.

Problems experienced by practitioners reported in the literature intrigued me. But I wondered if a more complete set of issues could be identified through a logical comparison of the ideals for naturalistic inquiry and standards for evaluation.

Logical Comparisons

In this section, criteria for judging the quality of naturalistic inquiries and evaluations are presented and contrasted. The criteria have been assembled from a variety of authoritative sources; but I do not claim these lists are exhaustive. However, they should be complete enough to diffuse the concern that "straw men" are being compared.

Evaluation Standards

Although a wide variety of models and methods for conducting evaluations have been proposed, nearly all evaluators agree that the purpose of evaluation is to assess the merit or worth of an evaluand. Evaluands can be multi-million dollar projects sponsored by governments or small course evaluations forming part of a curriculum evaluation system. Evaluations may be summative or formative, external or internal, requested or imposed.

In spite of all these variations, the Joint Committee on Standards for Educational Evaluation, appointed by twelve different national organizations (e.g., American Educational Research Association, American Psychological Association, American Personnel and Guidance Association, etc.), joined forces to identify a set of Standards for Evaluations of Educational Programs, Projects and Materials, (1981).

This set of standards appears to be the most comprehensive compilation of criteria against which an evaluation, educational or otherwise, could be judged. Brief summaries of each of the thirty standards as stated by the Joint Committee are reproduced below (by permission of Daniel Stufflebeam, the committee chair).

Utility Standards. The Utility Standards are intended to ensure that an evaluation will serve the practical information needs of given audiences.

1. Audience Identification - Audiences involved in or affected by the evaluation should be identified, so that their needs can be addressed.

2. Evaluator Credibility - The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum credibility and acceptance.

3. Information Scope and Selection - Information collected should be of such scope and selected in such ways as to address pertinent questions about the object of the evaluation and be responsive to the needs and interests of specified audiences.

4. Valuational Interpretation - The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for the value judgments are clear.

5. Report Clarity - The evaluation report should describe the object being evaluated and its context, and the purposes, procedures, and findings of the evaluation,

so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn and what recommendations were made.

6. Report Dissemination - Evaluation findings should be disseminated to clients and other right-to-know audiences, so that they can assess and use the findings.

7. Report Timeliness - Release of reports should be timely, so that audiences can best use the reported information.

8. Evaluation Impact - Evaluations should be planned and conducted in ways that encourage follow-through by members of the audiences.

Feasibility Standards. The Feasibility Standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal; they are:

9. Practical Procedures - The evaluation procedures should be practical, so that disruption is kept to a minimum and that needed information can be obtained.

10. Political Viability - The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

11. Cost Effectiveness - The evaluation should produce information of sufficient value to justify the resources expended.

Propriety Standards. The Propriety Standards are intended to ensure that an evaluation will be conducted legally, ethically and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results. These standards are:

12. Formal Obligation - Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

13. Conflict of Interest - Conflict of interest, frequently unavoidable, should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

14. Full and Frank Disclosure - Oral and written evaluation reports should be open, direct, and honest in their disclosure of pertintent findings, including the limitations of the evaluation.

15. Public's Right to Know - The formal parties to an evaluation should respect and assure the public's right to know, within the limits of other related principles and statutes, such as those dealing with public safety and the right to privacy.

16. Rights of Human Subjects - Evaluations should be designed and conducted, so that the rights and welfare of the human subjects are respected and protected.

17. Human Interactions - Evaluators should respect human dignity and worth in this interactions with other persons associated with an evaluation.

18. Balanced Reporting - The evaluation should be complete and fair in its presentation of strengths and weaknesses of the object under investigation, so that strengths can be built upon and problem areas addressed.

19. Balanced Reporting - The evaluator's allocation and expenditure of resources should reflect sound accounability procedures and otherwise be prudent and ethically responsible.

Accuracy Standards. The Accuracy Standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features of the object being studied that determine its worth or merit. These standards are:

20. Object Identification - The object of the evaluation (program, project, material) should be sufficiently examined, so that the form(s) of the object being considered in the evaluation can be clearly identified.

21. Context Analysis - The context in which the program, project, or material exists should be examined in enough detail, so that its likely influences on the object can be identified.

22. Described purposes and Procedures - The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

23. Defensible Information Sources - The sources of information should be described in enough detail, so that the adequacy of the information can be assessed.

24. Valid Measurement - The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the interpretation arrived at is valid for the given use.

25. Reliable Measurement - The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the information obtained is sufficiently reliable for the intended use.

26. Systematic Data Control - The data collected, processed, and reported in an evaluation should be reviewed and corrected, so that the results of the evaluation will not be flawed.

27. Analysis of Quantitative Information - Quantitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.

28. Analysis of Qualitative Information - Qualitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.

29. Justified Conclusions - The conclusions reached in an evaluation should be explicitly justified, so that the audiences can assess them.

30. Objective Reporting - The evaluation procedures should provide safeguards to protect the evaluation findings and reports against distortion by the personal feelings and biases of any party to the evaluation.

Criteria for Judging Naturalistic Inquiry

Like evaluation, the processes and standards for naturalistic inquiry have developed fairly independently in several different fields (e.g., anthropology, sociology, psychology, etc.). As of now, no cross-disciplinary committee has been formed to create an official statement of criteria. However, individuals (Guba,

1981; LaCompte and Goetz, 1982; Smith, 1979; Spindler, 1982; Williams, 1981; Wolcott, 1975) have proposed critical attributes of naturalistic studies which, taken as a whole, provide a reasonable set of principles against which the quality of a naturalistic study could be judged. No one inquirer would emphasize all these principles; in fact, most researchers and evaluators who use naturalistic methods would probably argue against using some subsets of these criteria in particular instances. However, this compilation does provide a working set of standards for use in the comparisons made in this paper.

1. There should be prolonged intensive engagement on site. As Wolcott (1975) said, "there must be time enough to meddle about and discover." The data gathering phase typically takes a year or more in many anthropological ethnographies. Although that length of time is not required for every study, the implication of this criterion is as Smith (1979) said, "you are there for all the variations so you are not misled." This criterion suggests not only lengthy engagement but also intensive engagement by the investigator. The researcher needs to be there frequently and long enough to penetrate the surface appearances.

2. Inquirers should ensure triangulation of sources, methods, variables, and instruments. This criterion suggests that the evaluator not rely on any one source of information or on one perspective. By collecting information through multiple sources and procedures and by focusing on multiple variables and concepts, chances are that the information obtained will be more credible than otherwise. This criterion also helps insure the external validity and reliability of the results of the study.

3. Reports should include presentation of "thick" raw description and contextual background. Understanding the object of study in its naturally occurring context is a hallmark of naturalistic inquiry and this criterion ensures the readers' ability to independently interpret and understand that object, be it a culture or an evaluand.

4. There should be no unaddressed internal contradictions in the report. This criterion ensures that naturalistic inquirers do not ignore or cover up contradictions in the data they collect. Rather, they should address those contraditions and try to incorporate them into the whole fabric of the data.

5. Natives (participants in the organizations or cultures being studied) should be able to review the written report and accept it as credible. If the people about whom the study is written disagree with a study, it certainly should not be disseminated for nonmembers of the culture to read as authentic.

6. The researcher should not have "gone native" nor should the researcher be an adversary of the object of study. In a sense this makes it sound like the researcher would be neutral and have no binding emotional ties. On the contrary this criterion suggests that the researcher should develop rapport with informants but not allow that rapport to develop to such an extent that the researcher becomes a member or an antagonistic criticizer of the object of study.

7. There should be evidence that issues have emerged during the study and analyses have been inductive. Although the researcher may begin the study with "foreshadowed problems" in mind, naturalistic studies should allow issues and the design to emerge. Analyses of data should be inductive and responsive to the contexts in which the data were gathered. There should also be evidence that participants have helped to define issues and hypotheses beyond the foreshadowed problems. Instruments used by the researcher should be created in the field based on informally gathered information rather than preordinately.

8. The native or emic perspective should be highlighted in the report. This criterion ensures that the author's pre-dispositions and biases do not determine the entire theme of the study. On the contrary, the "folk" perspective should be a central element of any naturalistic study.

9. The naturalistic inquirer should be both sensitve and perceptive, simultaneously sympathetic, skeptical, objective, and curious. In a very real sense the inquirer is the major collection instrument; therefore the person conducting the study should have those qualities that will ensure an insightful study.

10. Although statistical sampling is rarely used in naturalistic studies, sampling strategies certainly should be employed. The logic behind those sampling strategies should be clearly stated in reports. The strategies used should maximize the range of perspectives that are gathered by the naturalistic inquirer.

11. The report should provide the reader with a vicarious experience as a member in the setting studied. The reader should gain enough insight through reading a naturalistic report to be able to anticipate and interpret occurrances in the settings described therein almost as appropriately as do the members of the organizations or settings studied.

12. The naturalistic inquirer should document so completely the processes by which data are collected, samples are drawn, intepretations are made, etc., that an audit trail results. Independent researchers or auditors should be able to review the research activities, analyses and interpretations and make

independent judgments of the dependability and comfirmability of the claims made by the naturalistic inquirer.

13. It should be evident that negative instances were searched and accounted for by the naturalistic inquirer. This criterion helps protect the researcher against accusations that ideological beliefs were being supported through data collection as opposed to a dispassionate search for truth. By searching for evidence to counter his or her own data based conclusions, those conclusions can be modified to account for a more complete range of phemonema. Thus, the author of a naturalistic report becomes his or her own best critic.

14. Alternative interpretations of data should be seriously considered and evidence that these interpretations have been sought should be presented in a naturalistic study. This criterion ensures that naturalistic studies and their reports are not restricted to private interpretation but instead provide a rich enough data base that any reader can use the data to reach and check independent conclusions.

15. The inquirer's assumptions, predispositions, emotional reactions, and how those may relate to the inquir, should be well documented, either in the published report or in field notes. Certainly the presence of the inquirer should be noted in the report of findings. Since the inquirer is the principle instrument the reader needs to have some understanding of the perspectives used by the inquirer.

16. The inquiry should disturb as little as possible the processes and persons under study. The term "naturalistic" refers to non-contrived unobtrusive data collection. Of course, by having the participant observer or ethnographer present, there is some interruption of the natural processes. But the objective should be, through continual presence of the researcher on site to make that presence unobtrusive.

17. There should be evidence that participants are treated ethically. Precautions must ensure that confidences are not breached and that participants are treated, first as humans and only second as participants furthering the purposes of the research.

18. An important criterion by which all research and evaluation and naturalistic inquiry should be judged is whether or not it makes a clear contribution to knowledge and/or identifies insightful distinctions that may be used in solving problems.

19. A naturalistic inquiry must be well written. The writing should be of such a quality that readers can envision much of what the observers noted in their observations. Again,

the readers should have a vicarious experience and should almost feel as if they had been the researcher or a participant in the scene studied.

Comparison of Standards and Criteria

Table 1 summarizes a fairly simplistic comparison of the evaluation standards and the criteria for judging naturalistic inquiry. Whenever a criterion and a standard appeared to be complementary, a plus (+) was entered in the intersecting cell. Whenever they were neither complementary nor conflicting, the cell was left blank. Whenever it was judged that a naturalistic evaluator might encounter conflicts in attempting to meet both the standard and the criterion simultaneously, an identifying number was entered in the table.

Insert Table 1 About Here

At first glance, the most striking finding summarized in this table is the compatability of the standards and criteria. Of 570 potential conflicts, the standards and criteria are complementary in 242 (42%) instances. For example, when evaluation standard Number 18, "balanced reporting" is contrasted with the naturalistic criterion Number 8, "the emic or folk perspective is present", it is clear that including the participants' definitions, concerns, perspectives can do nothing to jeopardize the evaluation standard. On the contrary, meeting this naturalistic criterion enhances the balance of the report.

Likewise, 314 of the 570 (55%) comparisons are coded as blanks, signifying that the standards and criteria are neither complementary nor conflictual in a majority of instances. For example, evaluation standard Number 1, "Audience Identification" can be met or not met independent of whether naturalistic criterion Number 3, "thick, contextual description" is met or not.

In only 14 of the 570 comparisons (3%) were potential conflicts logically identified between the standards and criteria. And, as we shall see, even there the conflicts are not unavoidable, depending on the circumstances surrounding the study.

In the following section, each potential conflict will be discussed briefly so the standards and criteria most likely to clash in a naturalistic evaluation can be highlighted. Then, those contrasting principles will be used to reflect on my experience conducting a naturalistic evaluation.

The complementary and neutral interactions of the standards and criteria will not be emphasized in this paper. But clearly, a major conclusion to be drawn from the comparison of these two sets of principles is that they are highly compatible. The question raised is, "are the relatively few conflicts so powerful and insurmountable that they cancel out the benefits evidenced by the majority of non-conflicting interactions?"

Conflicts. Each of the fourteen conflicts identifed in Table 1 is described briefly below by number. Each evaluation standard is listed, followed by a slash and the conflicting naturalistic criterion.

1. Number 3--Information Scope and Selection/Number 7--Inductive/Emergent issues. Clearly there does not need to be a conflict between these two principles. As the "information scope and selection of standard" is stated, information collected should be responsive to the needs and interests of specified audiences. That same statement captures the essence of the naturalistic criterion, inductive emergent issues, which states that the information to be collected should not be restricted to a priori interests of the audiences, particularly of the persons commissioning and conducting the study. However, it is listed here as a potential conflict because as evaluation projects are commonly practiced, they can be interpreted to be meeting the information scope and selection standard when in reality they are not being responsive to the needs and interests of all specified audiences. In other words, audiences specified may be very restricted and specified in advance rather than allowing, in an emergent way, for the audiences relevant to the evaluand to identify themselves or be identified as the study progresses.

Likewise if a person begins a study with the idea that they will do it inductively and allow the audiences to emerge, they may find themselves in conflict with the sponsoring audience, the persons paying for the evaluation, who would prefer to identify specific audiences before the study begins and focus the attention of the study on those particular audiences and their needs and interests.

2. Number 6--Report Dissemination/Number 3--Thick Contextual Description. The potential conflict here is a practical one. Given extensive contextual description with many rich paragraphs containing direct quotations and detailed descriptions of the things observed during the evaluation, the number of pages in the report can be expanded considerably. If the report gets to be too long, its dissemination becomes problematic. Hence the evaluation standard stating that the evaluation findings should be disseminated to clients and other right-to-know audiences

so they can both assess and use the findings may be more difficult, the longer the evaluation report becomes. To meet the naturalistic criterion, length is less important than the richness of detail which will allow, in fact, the evaluation standard to be met more completely. That is, it will allow audiences to make their own assessments of the findings and to use them as they judge most appropriate.

3. Number 6--Report Dissemination/Number 12--Audit Trail Provided. This potential conflict is very similar to the one listed above in Number 2. The audit trail may be so huge that it will fill file cabinets and storage boxes. Clearly these materials cannot be duplicated and shipped to all members or potential audiences. However, the audit trail materials can be made available to individuals who would want to travel to the file bank of the evaluator to conduct an audit. So in principle there is no conflict here, in practice, there probably is a conflict.

4. Number 7--Report Timeliness/Number 1 Prolonged Engagement. This conflict seems to be the one that is emphasized in the literature most prevalently. Briefly stated, the conflict is that evaluators assume because of the evaluation standard that reports should be "timely" so that audiences can best use the reported information. The interpretation of timely most often is "as soon as possible." The assumption is that decision-makers which form at least part of the audiences would like to make decisions as fast as possible; and if the data from a study are not available when they are ready to make the decisions, then the study will not be useful to the decision-makers.

On the other hand, one of the most important naturalistic criteria is that of spending enough time on site to capture the essence of the object being studied as it passes through its various cycles. Clearly there is a high potential for conflict between these two principles. If an evaluator begins to study an object and discovers that it is changing over time and that the nature of the evaluand requires more time than has been allocated by the decision-making segments of the audience, then a compromise must be struck between these two principles. Either the investigator will spend the time to do the job right as stipulated by the naturalistic study or he/she will speed up the process in order to prepare a report that will be used by those particular members of the audience.

Clearly if the decision-makers are in accordance with the naturalistic criterion that a proper amount of time ought to be spent on site then they are not likely to require premature reporting. In other words, the seriousness of this potential conflict depends immensely on the audiences involved, their

requirements, the nature of the problem being studied, and so on.

5. Number 7--Report Timeliness/Number 12--Audit Trail Provided. This potential conflict is very similar to the previous two. If the evaluator must spend additional time compiling materials from which an audit may be conducted of the study, this takes time away from the actual conducting of the study and the writing of the report. Hence, the report may be submitted later than would be considered timely, either by the evaluator or by the audiences. Again, if the audiences are aware of the need to leave a trail for auditing purposes, they may be willing and in fact require that both the audit trail be provided and that the report be timely but with more realistic expectations in defining "timely."

6. Number 7--Report Timeliness/Number 13--Negative Instances Sought. Seeking negative instances to check the quality of the conclusions reached by an evaluator takes additional time and consequently may detract from the evaluation standard of report timeliness. Once again, if the audiences, particularly the sponsoring audiences who may be imposing restrictions of timeliness are aware of the value of a search for negative instances, they may be willing to allow a compromise between these two sets of principles. If they do not see the value of gathering negative instances or if constraints on them are such that they can not allow additional time, then a less than ideal compromise will have to be struck.

7. Number 9--Practical Procedures/Number 1--Prolonged Engagement. Similar to Number 4 above, the potential conflict here is that prolonged engagement on site may make the evaluation unfeasible, and impractical. A major strength of naturalistic inquiry is that over time the inquirer hopes to minimize disruption to the processes being studied. This is also one of the reasons for the "practical procedures" evaluation standard. But there is a risk that the presence of the evaluator could be disruptive and change the nature of the evaluand. Of course the naturalistic evaluator would be documenting changes as they were observed over time and it is assumed that through prolonged engagement, the observer would become a less obtrusive "part of the scene." But these precautions still might not overcome the problems of prolonged engagement.

The major concern in contrasting these two principles is that the requirement that the evaluator spend a considerable amount of time on site may compromise the realistic constraints of time, budget, staff and availability of participants which are intended to be guaranteed by the evaluation standard of practical procedures.

8. Number 9--Practical Procedures/Number 2--Triangulation. Again, the principle of triangulation requires that multiple data sources be tapped, at a variety of times, using a variety of procedures. All of these impose additional constraints on time, budget, and staff which may make the study impractical. On the other hand if the audiences, particularly the sponsoring agency is prepared for the additional cost and sees triangulation as beneficial, then there would not be a conflict between these two principles.

9. Number 12--Formal Obligations/Number 7--Inductive/ Emergent Issues. As in many of the other potential conflicts, if the audiences and parties to the evaluation can agree from the beginning that many of the issues will emerge as the study proceeds and they are willing to allow that open-ended nature to the design, then there is no conflict here. But many sponsors of evaluation and people who are involved with the evaluation want to know from the beginning of the study what is going to be done, how, by whom and so on. When a person insists on an emergent, inductive approach, it becomes difficult to forecast many of the activities and write them into a formal obligation statement.

10. Number 25--Reliable Measurement/Number 7--Inductive/ Emergent Issues. Reliable measurement depends on the opportunity to develop and refine an instrument or measure before it is used to collect data. When it is assumed that important issues will emerge as the data are gathered, there isn't as much time or interest in settling on a few issues to be built into such an instrument. It is conceivable that a naturalistic evaluator could let issues emerge naturally and then build those issues into instruments, taking time to establish the reliability of the measures. But it is not too likely, given the many other concerns a naturalistic study involves.

Naturalistic inquirers attempt to address the reliability issue through triangulation procedures, establishment of an audit trail, searching for negative instances and so on. They are less likely to take the time to create reliable instruments because that really needs to be done before the study is underway (as in a typical experiment, or a quasi experiment). The naturalistic inquirer is not able to create instruments <u>a priori</u> and still meet the criterion of allowing the participants and the data that are gathered to guide the study,defining issues as the study proceeds.

11. Number 25--Reliable Measurement/Number 8--Emic Perspective Highlighted. The naturalistic emphasis on the emic perspective suggests that the participants have an important role to play in defining the research issues in the study. That means the investigators cannot rely solely on pre-defined variables. Hence

the reliability of instruments to measure those variables will be very difficult to establish since the nature of the variables measured is not known until the study is under way.

This conflict is related to conflict number 10 because the emic perspective plays an important role in defining the emergent issues that are of such importance to naturalistic inquirers.

12. Number 25--Reliable Measurement/Number 16--Naturalistic, Unobtrusive, Uncontrived. The same point that was made in numbers 10, and 11 above holds true for this comparison. In order for the study to be naturalistic, unobtrusive, and uncontrived, the investigator cannot enter the field with a set of pre-formed or pre-determined instruments, however reliable they might be. Maintaining the naturalistic style, in most instances, prevents the researcher from developing instruments, establishing their realiability and then using them during a given study. Although on occasions this procedure might be possible.

13. Number 27--Analysis of Quantitative Information/Number 5--Natives Accept as Credible. Although quantitative information can be used very reasonably in a naturalistic inquiry, the gathering of it and the presentation of it can be problematic. The potential conflict here is between the nature of the quantitative information as it has been analyzed and presented and the ability and interest of the participants or natives in reading those analyses and considering them to be credible or not. Most participants in most settings in which naturalistic studies might be conducted are not likely to understand the reports of quantitative analyses. Therefore, they are not going to be able to consider those reports as being credible or truly reflective of the situation from their perspective. This does not mean, of course, that the quantitative information and the associated analyses are not valid or credible, it's just that the natives, as they would read reports which contain quantitative analysis are not likely to view these analyses as strengths or as true reflections of the characteristics of their environment.

14. Number 27--Analysis of Quantitative Information/Number 7--Inductive Emergent Issues. Usually the proper analysis of quantitative information necessitates systematic gathering of that information so that there are not many empty cells and so that the respondents to the instruments have an opportunity to answer all questions. An assumption underlying most quantitative analyses it is that instruments have been created to which all or a sample of all members of that particular population are exposed. In a naturalistic evaluation many of the issues which would form the basis for the creation of instruments would not be identified clearly before the study began. Therefore, fixed option questionnaire and test items could not be created

properly before the study began. Even if they were created *a priori* they would have to be modified as the study progressed. It is conceivable that toward the end of a naturalistic study an instrument could be generated based on the issues that emerged during the previous portion of the study, in which case these two principles would not be in conflict. But if the typical approach of bringing the instrument already created into the field before any other data are gathered were used, then these two principles would be in conflict.

Summary of Logical Comparisons. Looking across the evaluation standards to see which of them seem to clash the most with the criteria for judging naturalistic inquiry, it appears that the practicalities of conducting a naturalistic evaluation are the most troublesome. Report dissemination, report timeliness, and practical procedures account for seven of the fourteen potential conflicts. Also, the need for reliable instruments and the proper analysis of quantitative information account for nearly all the other potential conflicts (five of the fourteen).

Looking at the criteria for judging naturalistic inquiry that seem to conflict the most with the evaluation standards, the general openness or responsiveness of the naturalistic approach raises the most potential problems. The criteria of emphasizing inductive/emergent issues, the emic perspective and a naturalistic unobtrusive approach account for six of the fourteen conflicts. Likewise the impracticalities in terms of time for the evaluation are most clearly accounted for by the following criteria for judging naturalistic inquiry: prolonged engagement, triangulation, creation of an audit trail and searching for negative instances. These four criteria account for six of the fourteen opportunities for conflicts.

However, as the discussion above indicates, these potential conflicts do not appear to be unassailable. Given the proper circumstances, interests of audiences, resources, abilities of the person conducting the study, and so on, each of these conflicts can be resolved. In any given study not all conflicts are likely to be resolved but the nature of those resolutions and what brings them about raises interesting empirical questions.

In an effort to begin addressing some of those issues the following section of this report contains a brief analysis of a naturalistic evaluation which was conducted in a youth leadership class at Brigham Young University.

Of course, the author does not pretend to assume that this case is representative of all naturalistic evaluations that might be conducted or have been conducted. On the contrary, it is clear that this is a unique type of study as compared

to the large federally-funded, multi-site studies that have been analyzed and reported in the literature.

By reflecting on my experience in conducting one naturalistic evaluation and analyzing it in terms of the comparison of criteria presented in Table 1, I hope to make a much more confident judgment as to the importance of these potential conflicts between criteria for judging naturalistic studies and standards for judging the quality of evaluations.

Experiential Analysis

In this section, a brief description of a naturalistic evaluation is followed by an analysis of that experience in terms of the likely conflicts between naturalistic inquiry and program evaluation as identified by the logical comparison of evaluation standards and criteria for judging naturalistic inquiry, and the literature.

One Case of Naturalistic Evaluation

Context. Although our evaluation group had conducted several evaluations for the Physical Education College previously, none of the faculty or administrators had ever heard of participant observation or naturalistic inquiry methods. They were used to experimental design and survey research procedures that had been used previously. But the instructors of the course were the first to suggest that to evaluate their program adequately, we would have to participate in it.

For our part, we had never conducted a naturalistic evaluation and were a little hesitant to use all of our resources in conducting one. Therefore, one of our first steps was to interview the faculty members responsible for the course to ask them to help us begin to identify some of the evaluation issues for the project. Using a checklist of evaluation issues and questions to be considered in evaluating a course by Diamond and Sudweeks (1980), the two faculty members identified three major issues that were most important to them:

1. Course objectives (What are the objectives? What are the students' objectives in taking the course? How adequate, relevant and valuable are the objectives? How well are they being met according to students and staff as estimated by current methods used to grade the students?)

2. Instructional strategies (What are the instructional strategies, activities, presentations, materials, staff and student roles, etc.? How are they being used? How can the instructional strategies used in the course be improved to be more effective?)

3. Measuring progress (what are the procedures and criteria in judging students' progress and how well course objectives are reached? How could these procedures and criteria be improved? How can tests, assignments, and grading procedures be improved?)

We suggested to the faculty members that these three major issues would provide a place to begin formatively evaluating the course; but that we would be open to the identification of other, perhaps more relevant issues as we conducted the study. Although they agreed to this course of action, they soon decided that these three issues were the most important to them. They left the responsibility for identifying other issues to us, the evaluators.

In retrospect, I feel that the identification of these issues allowed us to begin doing something meaningful in terms of evaluation; but it made the study less naturalistic than I would want it if I were doing the study over. The checklist recommended by Diamond and Sudweeks is fairly comprehensive; but by focusing on those issues early, we did not allow as many issues to emerge naturally during the study.

After determining the major formative focus of the evaluation, we discussed with the instructors our intentions to participate in the class as students. They heartily agreed and recommended that the two evaluators split up so one could attend the course as taught by one instructor and the other could attend the other instructor's section. We agreed with this strategy, with the understanding that the two evaluators would meet regularly to discuss the separate experiences and to note similarities and differences.

A brief description of the course in which we participated will provide necessary background to understanding the nature of the evaluation experience. Youth Leadership is a course designed for recreation and youth leadership majors. The intended outcomes of the course were to teach the students leadership skills and provide them opportunities to practice skills as they prepare professionally and personally to be leaders. Class presentations, readings in the text, and small group presentations are the strategies used to teach these leadership principles. Class-wide and small group activites plus individual projects provide skills practicing experiences.

Several students who had previously completed the course were invited by the instructor to participate as "staff" members in the course. There were several staff functions which provided these students with leadership opportunities as they helped the instructors teach and manage the course. The instructors trained the staff and structured their activites by providing staff manuals which detailed the brief presentations the staff

were to make to students. The instructors conducted periodic interviews to help the staff set goals and review their progress. The instructors expected the staff to play a similar leadership role with the students, although the instructors also participated extensively in teaching, evaluation, and grading the students.

The one-semester (16 week) course consisted of a series of lessons on twenty-two leadership principles selected by the instructors. Early in the semester, all students participated in a campout for two nights and one and a half days. During that time, eight of the twenty-two principles were presented and special contests (called hurdles) were held between subgroups of the class. These contests were designed to give the groups an opportunity to practice some of the leadership principles. Another set of hurdles was held later in the semester covering other principles. Students were each given one opportunity during the semester to teach a short lesson to a subgroup of the class. They were required to compile a notebook of materials relevant to each of the twenty-two principles, and to set and complete one goal addressing each principle outside of class time.

Evaluation procedures. Four types of procedures were used in doing this evaluation. The major procedure, of course was participant observation. The instructors knew who the participating evaluators were and that they were playing the dual role of student and evaluator. The students and staff knew the evaluators only as fellow-students. Because these were college students and the evaluators were fairly young, it was not obvious that the evaluators were anything but regular students in these classes.

Throughout the semester, the participant observers took descriptive notes of their observations during class time and other course activities. They attended the campout and participated in hurdles, completed the regular assignments, and essentially behaved as regular students in the class. They summarized conversations they overheard and participated in with other students and staff members, recorded feelings they had both as students and evaluators and kept a running record of the events of the course. The observers also met privately with the instructors and gave them informal feedback throughout the semester regarding events and activities in which they participated. In this way, they provided formative evaluation results as results began to emerge.

Toward the end of the semester, a student questionnaire was developed based on the experiences of the participant observers and on questions raised by the instructors. Items on the questionnaire were designed to capture students' perceptions relative to the various evaluation questions and to validate the participant observers' suggestions for improving the course. Also, at the

end of the semester, a questionnaire was developed to be completed by the staff members.

Formal interviews were held with the instructors at the beginning, midway and at the end of the semester to give feedback to the instructors based on findings of the participant observers, to clarify evaluation issues, to help the evaluators better understand how the course was designed and what its purposes were, and to obtain the instructor's reactions to the ideas the evaluators were having about ways to improve the course.

The final report was prepared after the semester was over. There, the results that had been reported orally were described in much greater detail. Evidence from the participant observers' field notes and the questionnaire results were presented to support the conclusions and recommendations of the evaluation.

Although information had been collected on more issues than the three official evaluation objectives, the report itself focused almost exclusively on those concerns. As the semester proceeded, the evaluators came to realize that these were the issues of most importance to the instructors; so they wrote the report to respond to those interests. However, lengthy appendices were included describing the other issues of concern that were noted by the participant observers.

The evaluation activities evolved as the evaluation proceeded. It was our intention as participant observers to interview students and staff members more extensively and formally than we did. We became more realistic participants than we had originally intended, becoming very involved in the course as students and our role was cemented more as participants than observers. We had originally intended to meet more regularly with the instructors to give them feedback; but in accordance with our role as students, we did not meet with them except for three times: beginning, middle and end of the semester. One of the observers actually received a grade in the class. She completed all the assignments and authentically fulfilled the role of a graded student participant. As the other observer, I did not receive a grade in the class and did not complete <u>all</u> the assignments as I would have if I were taking the class for credit.

The two observers had intended to meet regularly to discuss the things they were learning through their observations; but we discovered that whenever we got together and talked about what we were seeing, we tended not to write as detailed and complete field notes. We found that once we talked to each other about the issues, we did not feel as much urgency to write about them. When we discovered that we were doing this, we stopped talking quite so much and spent more time writing and analyzing field notes.

We had intended to write a detailed portrayal of one "typical" student's experience in the course. As we compiled our notes, we realized the students' experiences had varied considerably and the instructors were mostly interested in the evaluation issues they had defined at the beginning of the semester. Therefore, in an appendix to the report, we summarized the field notes without writing a separate narrative portrayal.

The field notes provided the instructors with detailed evaluations, comments, feelings, and responses of the two observers as students and as evaluators. Comments were included on individual components of the course in the hope that they would be useful to the instructors in reforming the course, revising instructional materials and so on.

We decided these components of the report would be more valuable in a formative evaluation than a portrayal of what it would be like to be a student in the course for a third party who was not already involved in the course. In other words, the audience of the evaluation report became the two instructors and department administrators rather than potential students who might be considering taking the course or other faculty interested in offering leadership courses.

Analysis of Conflicts Between Criteria and Standards in Practice

In this section of the report, the fourteen potential conflicts described above will be reviewed and reflected upon in terms of my experience conducting the youth leadership naturalistic evaluation. The discussion will be organized around the same numeration system that was used above.

1. Number 3--Information Scope and Selection/Number 7--Inductive/Emergent Issues. We experienced this conflict in the youth leadership evaluation. We had intended to leave the specification of objectives open; but immediately upon presenting the checklist of potential evaluation objectives created by Diamond and Sudweeks (1980), our scope became very focused, much as the evaluation standard Number 3 suggests that it ought to be.

We responded to the needs and interests of the faculty because they mentioned three of those objectives and indicated they would like us to focus our efforts there. As a result, we limited ourselves pretty much to those three objectives. We did note other important issues that ought to be considered in improving this course as we participated in it as students but the entire tenor of the study was colored heavily by the initial focus. This happened to such an extent that we did

not adequately meet the naturalistic criterion of allowing issues to emerge naturally.

2. Number 6--Report Dissemination/Number 3--Thick-Contextual Description. Although we gathered considerable description, rich in detail and contextual information, most of that description did not find its way into the report. We included excerpts from our field notes in the appendices of the report but did not do the detailed portrayal of one student we had planned because we found our report was going to be read only by the instructors. They were not as interested in a detailed portrayal as they were in the results of the survey that we designed and our overall recommendations for improving the course.

If the report could have been reasonably disseminated to future students of the course or to other faculty who were designing similar courses, than the rich description might have played a more important role and we probably would have included it more extensively in our report.

3. Number 6--Report Dissemination/Number 12--Audit Trail. We clearly did not plan to disseminate all the materials that would provide an audit trail. There was too much information to do so. Some outline of an audit trail is apparent in the appendix but we did not meet the naturalistic inquiry criterion of providing a detailed audit trail as well as we could have done. On the other hand we disseminated the report to the audiences for whom it seemed most applicable given that this was a formative evaluation.

Perhaps if we had been held accountable publicly to provide an audit trail we would have placed more emphasis on it; but we found that we wanted to be responsive mainly to the instructors, providing them with formative information. Their interests were focused mainly on the information we were providing, not on establishing an audit trail. So we sacrificed the criterion for the evaluation standard.

4. Number 7--Report Timeliness/Number 1--Prolonged Engagement. Clearly there was a conflict between report timeliness and prolonged engagement on site. We found that we spent so much time acting like students participating in student activities and discussions in the class that we had little or no time to conduct the analysis and writing activities that we should have been conducting all along during the semester. So when the semester came to an end, we had to get busy doing the bulk of the analysis and writing, which postponed the submission of the report. We tried to compensate for this conflict by providing oral reports to the instructors throughout the semester about findings that we were encountering. We raised issues and suggested potential modifications they could make to the course. The

instructors were more responsive to those oral reports than to the final written report which was submitted long after the semester ended. So in this instance the evaluation standard was subordinated to the naturalistic inquiry standard.

5. Number 7--Report Timeliness/Number 12--Audit Trail. The need for timeliness of the report did present a conflict with the criterion that we were to establish an audit trail. However, it didn't seem to be a real problem in this particular case because our principal audience, the instructors, were willing to accept our oral reports and suggestions throughout the study and they were not anxious to have the evidence of an audit trail to judge the credibility of our suggestions. They compared what we said to their own experience in the course and the financial and time restrictions imposed on them for improving the course.

The experience we had here suggests that the establishment of an audit trail is more important to methodologists and outside reviewers than it is to the participants. In an evaluation, particularly in a formative evaluation, the persons responsible for improving the product are the principal audience to which an evaluator should respond. Hence our minimal emphasis on accumulating audit trail information was probably the correct action. The fact that a well defined audit trail was not made available, was not a drawback to the quality of the study as far as the local participants were concerned.

6. Number 7--Report Timeliness/Number 13--Negative Instances Sought. There was a conflict between report timeliness and the requirement that we seek for negative instances. A majority of the study time was spent as participant observers developing hypotheses about ways to improve the course. We developed a questionnaire to organize these hypotheses and then submitted the questionnaire to the students to see if they would confirm or disconfirm our ideas about the best ways to improve the course.

We submitted this questionnaire to the students at the end of the semester though and did not have time then to pursue the issues further with students who responded differently than we predicted they would. The questionnaire embodied our understanding of what was going on but we were not ready to administer it until the end of the semester and then there was no time to follow up. We could have pursued it the next semester but felt pressure to write the report and finish the study instead.

7. Number 9--Practical Procedures/Number 1--Prolonged Engagement. There was a potential conflict between the use of practical procedures and prolonged engagement. In retrospect we could have spent much longer studying these courses with

other sets of students; however, the instructors expected answers by the end of the semester and they didn't really anticipate that we would want to continue our involvement. Also the college wasn't prepared to spend more of their budget on this particular course so we were obliged to stop the study when the semester ended.

Our willingness to not continue the study into the following semesters was a response to the evaluation standard of sticking with practical procedures which compromised the naturalistic inquiry criterion of prolonged engagement. However, we did prolong our engagement much more than had been done in previous studies for this college. To establish an evaluator's presence daily in a class over an entire semester was unprecedented. So in terms of tradition, one could almost argue that we compromised the evaluation standard more than we did the naturalistic inquiry criterion.

8. Number 9--Practical Procedures/Number2--Triangulation. Triangulation was used in this study. We conducted participant observations, interviews and administered surveys to two different sub-populations of students and the staff members. But clearly, triangulation is a principle that can be expanded and used much more extensively than we did here. To the degree that we restricted our triangulation procedure there was a compromise between triangulation and the use of practical procedures. However, that compromise was not difficult to reach in this case.

9. Number 12--Formal Obligations/Number 7--Inductive/Emergent Issues. We formally agreed with the faculty to address three objectives that they had identified at the beginning as being critical. We also told them that we were going to leave the study open so that other issues that might be of importance could emerge. However, we ended up focusing mainly on those three objectives.

I don't view this as a compromise between these two principles. The faculty were willing for us to include emergent issues, agreeing to it in the formal obligation. The reasons we focused our attention on their three objectives were more a function of time pressures and our desire to be responsive to the principal audiences than to a conflict between the formal obligation standard and the criterion to have inductive emergent issues play a role in the study.

10, 11, 12. The conflicts described as numbers 10, 11, and 12 all have to do with the evaluation standard of reliable measurement. The associated naturalistic inquiry criteria are: inductive/emergent issues, emic perspective highlighted, and naturalistic, unobtrusive procedures used. In this case the naturalistic inquiry criteria were emphasized over the need

to have reliable instruments. However, many of the other criteria for judging naturalistic inquiry are supportive of the effort to make the study be reliable or dependable as was clarified earlier in this paper and is demonstrated by the number of pluses in Table 1 for line 25, reliable measurement.

In this study a questionnaire was created as issues emerged throughout the semester to include not only the three major issues but other issues that were raised by the participant observers. By the time the instrument was created, the semester was almost over and so it was administered without elaborate procedures for establishing its reliability.

Likewise, the emic perspective was highlighted over the reliability of the instrument by our emphasis on identifying issues in students' terms to use on the questionnaire. We did not impose our perspective by creating a questionnaire a priori with high reliability.

Finally, we chose to emphasize the unobtrusive approach by not entering the classrooms at the beginning of the semester to collect information that would help establish the reliability of a pre-fashioned instrument. In fact we tried to make the instrument as unobtrusive as possible by administering it at the end of the semester when it was too late to establish its reliability and refine it.

13. Number 27--Analysis of Quantitative Information/Number 5--Natives Accept as Credible. This potential conflict was not a real problem in this case. The instructors valued the quantitative data that were provided almost more than they valued the qualitative descriptions. Our reports of the survey results were considered credible by them, even more so than several of our participant observation reports.

The instructors came to view us as students and so they were anxious to see how the rest of the students had felt about the recommendations we were making and did not want to rely solely on our qualitative analysis. In this case, the natives (the instructors) viewed the statistical summaries that we made of our survey as important additions, enhancing the credibility of the reports.

14. Number 27--Analysis of Quantitative Information/Number 7--Inductive, Emergent Issues. There was a small amount of compromise between these two criteria in our case. Much more powerful quantitative analyses would have been possible if additional quantitative instruments had been used. We built the questionnaire as the study progressed as opposed to testing a priori hypothesis but we were able to use quantitative items to gather students' responses to these emergent issues. The

proper quantitative analysis procedures were used given our purpose. There was no serious conflict except we did not have time to properly refine the instrument which is an important requirement in allowing proper analysis of quantitative data.

Analysis of this case in light of problems identified in the literature. Most of the problems identified by the literature which other individuals have encountered in applying naturalistic inquiry methods to evaluation and policy making studies have been accounted for by the logical analysis between evaluation standards and criteria for judging naturalistic inquiries. That is to say, potential problems from the literature listed on pages 3-4 of this report as numbers 1, 2, 4, 5, 6, 9 and 10 have been accounted for. However, problems 3, 7 and 8 were not
adequately addressed by the logical comparison of criteria and standards. Therefore, I would like to briefly analyze the naturalistic
study which we conducted in light of those three problem areas.

1. Problem Number 3 states that there is a potential conflict between the emphasis on description in naturalistic studies and the emphasis on judgment in evaluation studies. We found in this particular case that the audience wanted our judgments; so we gave them along with associated descriptive information. We did not provide a lot of pure description. However, in reflecting on the idea of description, I'm not sure that that is really an ideal of naturalistic inquiry. One of the criteria for judging naturalistic inquiry. (Number 15) states that the inquirers' presence and biases should be documented, suggesting that naturalistic inquirers recognize that they have biases and that they will make judgments and interpretations.

Wolcott (1975) stated that he tries to withhold judgment while he is collecting the data and then uses his judgments and interpretations in developing his written report. In other words, judgment has a role to play in naturalistic inquiry. It is not to be avoided completely.

Although the emphasis on judgment in evaluation may be much higher than it is in naturalistic inquiry, a useful object in either case is to create a report that will help the reader make his/her own judgments. To do this, rich case descriptions, intepretations and the biases of the inquirer should be clearly stated in the report.

I felt that by including our thoughts and feelings throughout the study in the appendices that the readers of our evaluation report would be able to recognize more clearly why we made the

value judgments, collected the information and made the recommendations we did. I did not find a conflict in this case.

2. Problem Number 7 suggests the evaluator must mediate conflicting roles and interests between the requirements of the evaluation sponsors and the interests of participants. It was true in our case that we did not give equal weight to the interests of faculty and students. On the contrary, we were there to help the faculty improve the course, which we and they hoped would eventually improve the situation for students.

We attempted to respond to two extreme perspectives: that of the faculty by addressing some of their objectives and of the students by assuming the role of students during the course of the study. By playing both of those roles simultaneously we were able, we believe, to strike a fairly useful balance between sometimes conflicting needs and expectations of these two groups. However, we acknowledge that we did emphasize more formally (in the report) the faculties' interests. In any naturalistic study, be it evaluative or not, the inquirer must strike a balance in addressing conflicting roles and interests.

3. Problem Number 8 suggests there is a potential conflict between the emphasis on understanding in naturalistic inquiry and an emphasis on evaluating. This conflict identified in the literature is very similar to the one listed above between description and judgment. I believe, after conducting the youth leadership evaluation that there is a potential for serious conflict here. By focusing our efforts on the act of collecting information to improve the course and to make judgments about the merit of the course, we were unable to spend the time and effort it would take to truly understand the positions of the various parties involved in the course. Students' motives and activities were tapped only as they related to the evaluation objectives. And true understanding as it is reflected by the term "verstehen" was probably not obtained in this evaluation.

Summary of Experiential Analaysis. Reviewing all fourteen of the potential conflicts identified in the logical analysis (as summarzied in Table 2), it appears that for five of the potential conflicts (5, 8, 9, 13, and 14) there was no serious conflict in this particular case. In four of them (4, 10, 11, and 12) there was a conflict and the naturalistic inquiry criterion was supported over the evaluation standard. In five of the potential conflicts (1, 2, 3, 6 and 7) the evaluation standard was supported over the naturalistic inquiry criterion. Summarizing the additional three conflicts described in the literature, we experienced two of three conflicts in this case. That makes a total of 11 of the 17 potential conflicts which were conflicts in this naturalistic evaluation. Of course in

all of these conflicts, there was compromise struck between the standards, and criteria with neither the standards nor the criteria receiving exclusive emphasis.

Conclusions

As this analysis indicates, there is compromise in attempting to jointly conduct an evaluation and to use naturalistic inquiry procedures. In fact, some of the naturalistic inquiry criteria are not met as completely as they could be. Of course this is the case in all naturalistic inquiry. Not all the criteria are equally addressed in every instance in which someone has applied the naturalistic paradigm for either research or evaluation purposes.

Depending on the needs of the evaluation client, the other evaluation audiences involved, the resources available, the interests of the inquirer, the time constraints, the nature of the evaluand and several other variables, the power of the naturalistic paradigm can be diluted somewhat when applied to evaluation; but the overall quality of the evaluation can also be substantially improved.

The advantages for making value judgments in terms of clearly understanding the nature of the evaluand, discovering unintended outcomes and allowing more of the emic perspective and interests of the participants to guide the evaluation responsibly, justifies the necessary compromises, especially given the number of complementary interactions as summarized in Table 1. The information needs will often warrant the use of a naturalistic approach to evaluation.

The implication of this conclusion is that naturalistic techniques should neither be avoided nor embraced thoughtlessly in evaluation projects. Rather when the information needs, the audiences involved, and other impinging variables so dictate, the naturalistic paradigm can be a very useful tool in assessing merit or worth through the evaluation process.

Table 1

Comparison of Evaluation Standards and Criteria for Judging Naturalistic Inquiry

Evaluation Standards	1 Prolonged Engagement	2 Triangulation	3 Thick Contextual Description	4 No Internal Contradictions	5 Natives Accept as Credible	6 Researcher not "native"	7 Inductive, emergent issues	8 Emic Perspective Highlighted	9 Sensitive, Perceptive, Inquirer	10 Logical Sampling Strategy	11 Vicarious Membership for Reader	12 Audit Trail Provided	13 Negative Instances Sought	14 Alternative Interpretations Sought	15 Inquirer's Qualities Documented	16 Naturalistic, Unobtrusive	17 Ethical Treatment	18 Clear Contribution	19 Well Written	
1. Audience Identification	+	+			+		+	+		+				+				+		1
2. Evaluator Credibility	+			+	+	+			+			+	+	+	+			+	+	2
3. Information Scope Selection	+	+			+	+	1	+	+	+		+	+	+	+			+		3
4. Valuational interpretation			+	+		+						+		+	+					4
5. Report Clarity			+									+							+	5
6. Report Dissemination			2										3							6
7. Report Timeliness	4											5	6							7
8. Evaluation Impact	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+		+	+	8
9. Practical Procedures	7	8	+				+		+	+						+	+			9
10. Political Viability	+	+			+	+	+	+	+	+		+	+	+	+	+	+	+		10
11. Cost Effectiveness																		+		11
12. Formal Obligation							9										+			12
13. Conflict of Interest	+	+	+		+	+	+	+	+	+		+	+	+	+		+			13
14. Full and Fran Disclosure			+			+		+				+	+	+	+				+	14
15. Public's Right to Know																	+			15
16. Rights & Human Subjects																	+			16
17. Human Interactions					+	+	+	+						+		+	+			17
18. Balanced Reporting		+	+	+	+	+		+	+		+	+	+	+	+			+	+	18
19. Fiscal Responsibility												+						+		19
20. Object Identification	+	+	+		+		+	+	+		+							+	+	20
21. Content Analysis	+	+	+		+		+	+	+		+		+	+	+	+			+	21
22. Described Purposes and Procedures			+							+		+			+					22
23. Defensible Information Sources		+	+							+		+	+		+					23
24. Valid Measurement	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+		24
25. Reliable Measurement	+	+	+			+	10	11	+	+		+	+	+	+	12				25
26. Systematic Data Control	+	+	+	+	+							+	+							26
27. Analysis of Quantitative Information			+	+	13		14			+		+		+						27
28. Analysis of Qualitative Information	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	28
29. Justified Conclusions		+	+	+	+	+	+	+		+	+	+	+	+	+				+	29
30. Objective Reporting			+	+	+	+		+		+	+	+		+	+				+	30
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	

+ = Standards and Criteria complementary.
#'s 1-14 = Standards and Criteria potentially conflictual.
Blank = Standards and Criteria neither complementary nor conflictual.

Table 2
Summary of Which Potential Conflicts and Which Criteria and Standards were Emphasized in One Case of Naturalistic Evaluation

	Criteria for Judging Naturalistic Inquiry							
Evaluation Standards	Prolonged Engagement 1	Triangulation 2	Thick Contextual Description 3	Natives Accept as Credible 5	Inductive, Emergent Issues 7	Audit Trail Provided 12	Negative Instances Sought 13	Naturalistic, Unobtrusive 16
3. Information scope & selection					1.E			
6. Report Discimination			2.E			3.E		
7. Report Timeliness	4.N					5.B	6.E	
9. Practical Procedures	7.E	8.B						
12. Formal Obligation			9.B					
25. Reliable Measurement			10.N		11.N			12.N
27. Analysis of Quantitative Information				13.B	14.B			

Key

#'s in cells refer to conflict numbers as listed in text

E = Evaluation standard was emphasized over the naturalistic inquiry criterion.

N = Naturalistic inquiry criterion was emphasized over the evaluation standard.

B = The evaluation standard and the naturalistic inquiry criterion were both emphasized. There was no true conflict.

REFERENCES

Diamond, R. M., & Sudweeks, R. R. (1980). A comprehensive approach to course evaluation. Journal of Instructional Development, 29(2), 75-91.

Fetterman, D. M. (1983). Guilty knowledge, dirty hands, and other ethical dilemmas: The hazards of contract research. Human Organization, 4293), 214-223.

Fetterman, D. M. (1982). Ethnography in education research: The dynamics of diffusion. Educational Researcher. 11(2), 17-22.

Guba, E.G. (1981). Criteria for assessing thetrustworthiness of naturalistic inquiries. Educational Communication and Technology Journal, 29(2), 75-91.

Guba, E.G. & Lincoln, Y.S. (1981), Effective Evaluation: improving the usefulness of evaluation results through responsive and naturalistic approaches. San Francisco, Calif.: Jassey-Bass Publishers.

Guba, E.G. (1978). Toward a methodology of naturalistic inquiry in educational evaluation. CSE Monograph Series in Evaluation. Los Angeles, CA: Center for the Study of Evaluation, UCLA Graduate School of Education, University of California, Los Angeles, 8.

Joint Committee on Standards for Educational Evaluation (1981). Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill Book Company.

LaCompte, M. D., & Goetz, J. P. Problems of reliability and validity in ethnographic research. Review of Educational Research, 52(1), 31-60.

Patton, M.Q. (1980). Qualitative Evaluation Methods. Beverely Hills, CA: Sage Publications.

Smith, A. G., & Robbins, A. E. (1982). Structured ethnography: a study of parental involvement. American Behavioral Scientist, 26(1).

Smith, L. M. (1979). An evolving logic of participant observation, educational ethnography, and other case studies. L. Schulman (ed). Review of Research in Education, (Vol. 6), Chicago, Illinois: Peacock Press.

Spindler, G. (1982). Doing the Ethnography of Schooling: Educational Anthropology in Action. (General introduction, G. Spindler (ed.). New York: Holt, Rhinehart and Winston.

Stake, R.E. (1975). Program evaluation, particularly responsive evaluation. Occasional Paper Series, Kalamazoo, Michgian: The Evaluation Center, Western Michigan University, 5.

Stake, R.E. & Easley, J. (1978). Case Studies in Science Education. Urbana, Illinois: University of Illinois.

Williams, D. D. (1981). Understanding the work of naturalistic researchers. (Unpublished dissertation).

Wolcott, H.F. (1975). Criteria for an ethnographic approach to research in schools. Human Organization, 34(2), 111-127.

Wolcott, H.F. (1982). Mirrors, models and monitors: educator adaptations of the ethonographic innovation. In Doing the Ethnography of Schooling, Spindler, G. (ed.). New York: Holt, Rinehart and Winston.